

Forecasting Consistency and the Small Area Secondary Variables Allocation Model (SASVAM)

Simon Choi and SungHo Ryu***

* Southern California Association of Governments

** University of Southern California

Presented at the 2008 COG/MPO Mini Conference,
San Diego, California, August 8-9, 2008



Presentation Outline

USC

- Background
- Forecasting Approach
- SASVAM
- Challenges
- Future Directions



Background

- ❑ 2008 Regional Transportation Plan (adopted in May 2008)
- ❑ Growing demand of socioeconomic variables for RTP & EJ analysis.
- ❑ Difficulty
 - Many socioeconomic variables (linkage of 100+ variables)
 - Small area (grid cell, parcel)
 - Long term (30 years)

5

Background

- ❑ Two Step Approach
 - Major “popular” variables
 - ❑ Population
 - ❑ Households
 - ❑ employment

6

Background

- Two Step Approach
 - Secondary “derived” variables
 - Population: age, race/ethnicity.
 - Households: number of persons, number of workers, poverty level, age of householders, race/ethnicity and income level of householders.
 - Workers: income level.
 - Employment: industry sectors, income level.

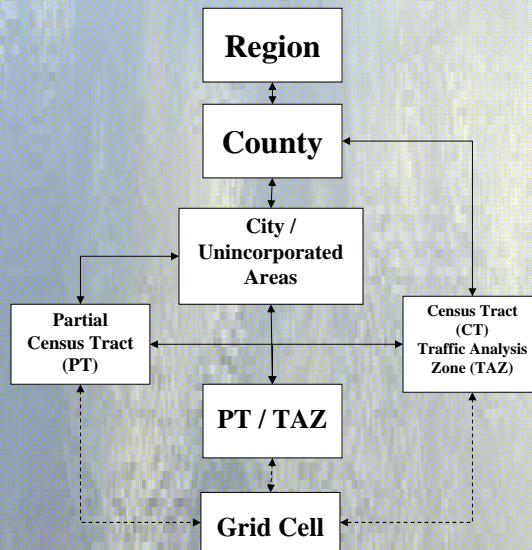
7

Major vs. Secondary

- Major
 - Region – County – City – TAZ- CT/TAZ – Grid Cell (5.5 acres=150m*150m)
 - Communicative process
- Secondary
 - Region – County – TAZ
 - Consistency

8

Forecasting Geography



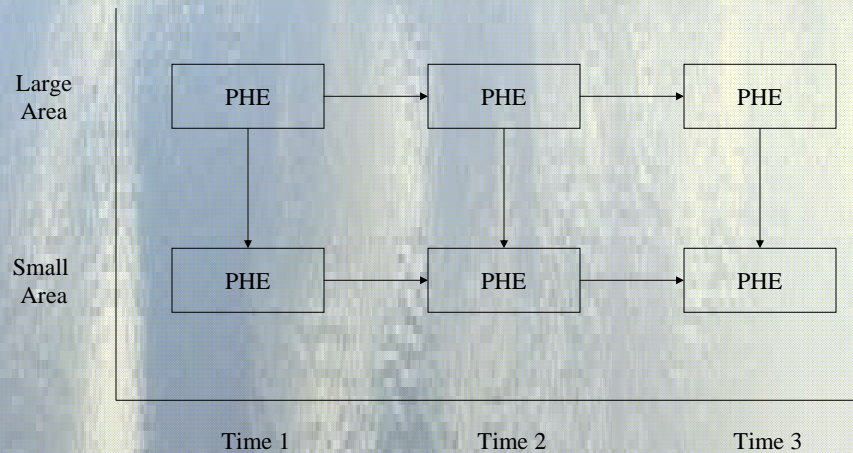
9

Forecasting Approach

- ☐ Reasonable
- ☐ Consistent
- ☐ Acceptable
- ☐ Collaborative

10

Consistency (major variables)



11

Consistency (major variables)

- Linkage of major variables
 - Population and Households (household formation, average household size)
 - Population and Employment (labor force participation rate, P/E ratio)
- Spatial
 - Disaggregation vs. Aggregation
- Temporal
 - Monotonous vs. Fluctuation

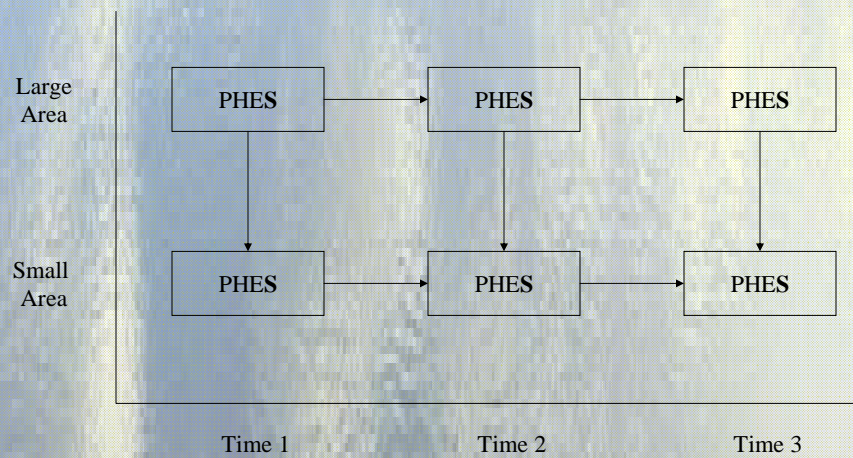
12

Consistency

- Allow socioeconomic variables to communicate...

13

Consistency (secondary variables)



14

Consistency (secondary variables)

- ❑ Linkage of major variables and secondary variables-distribution probability
 - Population: age distribution, race/ethnic distribution
 - Households: age distribution, race/ethnic distribution, household size distribution, worker size distribution, income distribution
 - Workers: income distribution
 - Employment: industry sector distribution, income distribution
- ❑ Spatial: Disaggregation vs. Aggregation
- ❑ Temporal: Monotonous vs. Fluctuation

15

SASVAM

*Small Area
Secondary Variables
Allocation Model*

16

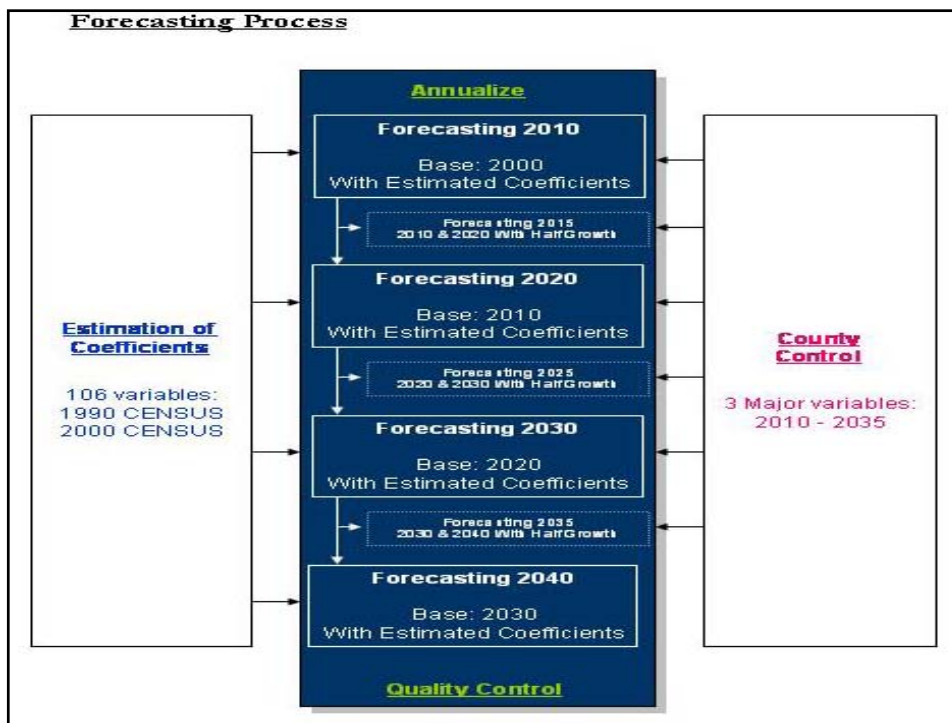
Objectives of SASVAM

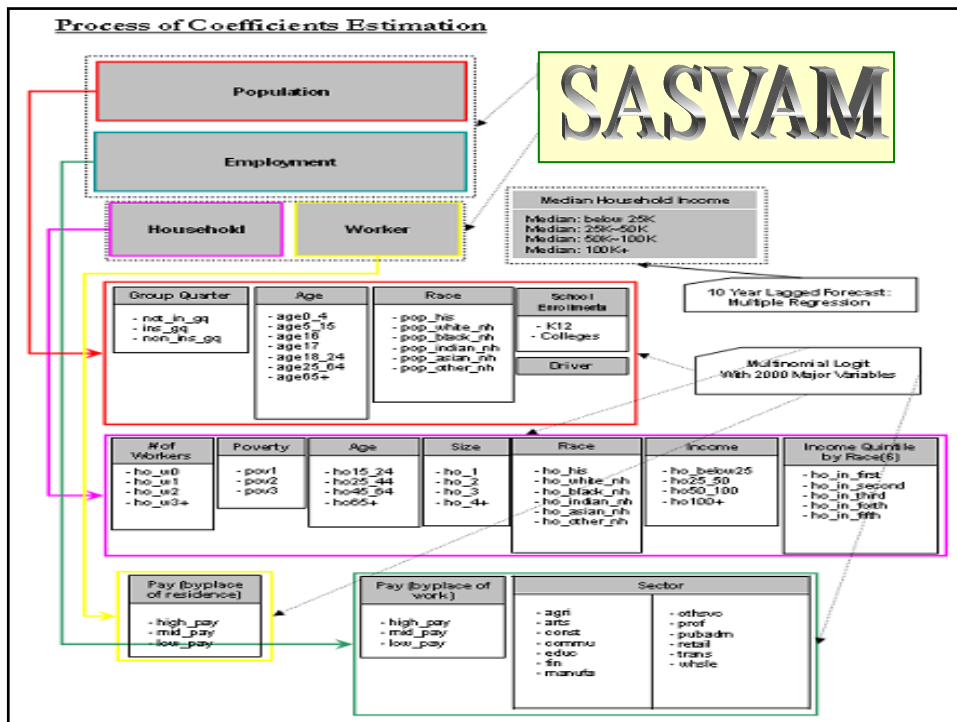
- Forecasting Reasonable, Consistent, Acceptable, and Collaborative Small Area Data Set,


Considering:

- 1) relationship between major var. and secondary var.
- 2) geographical relationship
- 3) temporal relationship

17








SOUTHERN CALIFORNIA
ASSOCIATION OF
GOVERNMENTS

Example: Estimation of Coefficient (Multi-Nomial Logit Model)



USC

□ **Population by Race/Ethnicity**

$$Race_{2000} = \beta_0 + \beta_1 Race_{1990} + \beta X_{1990} + \varepsilon$$

Race/Ethnicity(2000) = Hispanic pop(1990) + NH White(1990) +
 NH Black(1990) + NH Indian(1990) +
 NH Asian(1990) + NH Other(1990) +
 Age 25-64(1990) + Employment(1990)
 Median HH income(1990) +
 Pop density(1990)

Pseudo R-Square : 0.4407

**Data Source: Census Bureau SF3 Census Tract data (1990 & 2000)
CTPP (1990 & 2000)**

20

Initial Forecasting Process

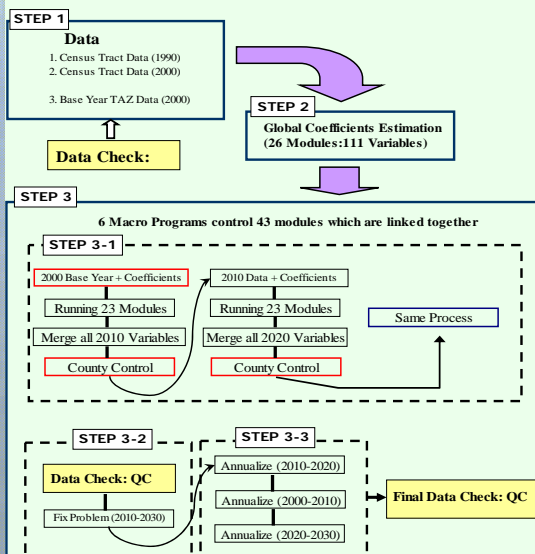
- Estimate Coefficient (1990 & 2000)
- Estimate future distribution of each secondary variable with *Multi-nomial Logit Regression*

$$\text{Pr ob}(y = j) = \frac{\sum_{k=1}^k \beta_{jk} x_k}{1 + \sum_{j=1}^{J-1} e^{\sum_{k=1}^k \beta_{jk} x_k}}$$

- 2010 distribution = Coefficient * 2000(TAZ)
- 2010 Race = 2010 population * 2010 distribution

21

SASVAM Process



- 23 secondary var. forecasting modules
- 18 quality control modules
- Model running time...

22

Model Strengths (I)

- Reflect historical trend (1990-2000): Coefficient Estimation
- Constrained by higher level of geography: County control (TAZ level) or TAZ control (Grid cell level)
- Consistency between major variables and secondary variables
 - If Pop=0 then All pop related Var.=0
 - Residential pop (not in GQ) > Worker
 - Residential pop (not in GQ) > Household
 - If Residential pop=0 then HH=0
 - Age_16+ > Worker

23

Model Strengths (II)

- Consistency between secondary variables

HH by # person vs. HH by # worker

HH(10): H_1(3), H_2(4), H_3(2), H_4_over(1)

HH(10): Hw0(1), Hw1(4), Hw2(3), Hw3+(2)

$$H_w3+(2) < H_3(2) + H_4_over(1)$$

- Check 250 rule of thumbs
- Temporal consistency: Annualize data after forecasting decade data

24

Model Challenges

- Importance of major variables
 - Population (or Households) = 0, 1 or 2 (abnormal distribution).
- Group quarter population
- Integer vs. Decimal
 - Calculation (Rounding Issue)
 - Adjustment Process
 - Reality

25

Future Directions

- Grid cell or parcel level projection
- Add more secondary variables
 - Housing by type (SF, MF)
 - Households by tenure (Owner, Renter)
- Refine coefficient estimation based on theoretical background
- More integrated forecasting process into other models (land use, transportation models)

26

Thank You

- Simon Choi: **choi@scag.ca.gov**
- SungHo Ryu: **sungryu@usc.edu**